

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2002, 11:15:22 ; Search time 30.05 Seconds  
(without alignments)  
66.533 Million cell updates/sec

Title: US-09-394-019A-248  
Perfect score: 96  
Sequence: 1 KDPXGLEHGDINGXPKGY 18

Scoring table: BLOSUM62  
Gap 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

1:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					Description
Result No.	Score	Query Match	Length	ID	
1	92	95.8	18	22	Protease indicator
2	90	93.8	18	22	Protease binding s
3	86	89.6	18	22	Protease indicator
4	84	87.5	18	22	Protease binding s
5	82	85.4	18	22	Protease binding s
6	80	83.3	18	22	Protease indicator
7	80	83.3	18	22	Protease binding s
8	78	81.2	18	22	Protease indicator
9	73	76.0	18	19	Protease binding s
10	73	76.0	18	22	Fluorogenic protea
11	73	76.0	18	22	Protease indicator

12	71	74.0	18	22	AAG73082	Protease binding s
13	71	74.0	18	22	AAG73086	Protease binding s
14	71	74.0	18	22	AAG73090	Protease binding s
15	71	74.0	18	22	AAG73091	Protease binding s
16	70	72.9	18	22	AAG73318	Protease indicator
17	68	70.8	18	22	AAG73088	Protease binding s
18	65.5	68.2	19	22	AAG73302	Protease indicator
19	65	67.7	18	19	AAW82120	Fluorogenic protea
20	65	67.7	18	22	AAG73319	Protease indicator
21	64	66.7	18	19	AAW82119	Fluorogenic protea
22	64	66.7	18	19	AAW82115	Fluorogenic protea
23	63.5	66.1	17	22	AAG73303	Protease indicator
24	63.5	66.1	17	22	AAG73304	Protease binding s
25	63.5	66.1	19	22	AAG73084	Protease binding s
26	63	65.6	18	22	AAG73089	Protease binding s
27	63	65.6	18	22	AAG73099	Protease binding s
28	62.5	65.1	17	22	AAG73085	Protease binding s
29	62	64.6	18	22	AAG73094	Protease binding s
30	62	64.6	18	22	AAG73098	Protease binding s
31	61	63.5	18	19	AAW82136	Protease indicator
32	60	62.5	16	22	AAG73301	Protease indicator
33	60	62.5	18	19	AAW82118	Fluorogenic protea
34	60	62.5	18	19	AAW82122	Fluorogenic protea
35	59	61.5	18	19	AAW82121	Fluorogenic protea
36	59	61.5	18	19	AAW82110	Fluorogenic protea
37	59	61.5	18	19	AAW82113	Fluorogenic protea
38	59	61.5	18	19	AAW82114	Fluorogenic protea
39	59	61.5	18	19	AAW82238	Fluorogenic protea
40	59	61.5	18	19	AAW82240	Fluorogenic protea
41	59	61.5	18	22	AAG73079	Protease binding s
42	59	61.5	18	22	AAG73080	Protease binding s
43	59	61.5	18	22	AAG73117	Protease binding s
44	59	61.5	18	22	AAG73118	Protease binding s
45	59	61.5	18	22	AAG73258	Protease indicator

ALIGNMENTS

RESULT	1
ID	AAG73324
AC	AAG73324 standard; Peptide; 18 AA.
AC	AAG73324;
DT	14-AUG-2001 (first entry)
XX	Protease indicator compound peptide #53.
DE	Protease detection; peptide cleavage; enzyme activity; fluorogenic;
KW	viral infection; cancer metastasis; emphysema; arthritis;
KW	thrombosis; haemophilia.
XX	Synthetic.
OS	
XX	
FH	Key
FT	Modified-site 4
FT	Location/Qualifiers
FT	/label= OTHER
FT	/note= "designated J in the specification"
FT	14
FT	Modified-site
FT	/label= OTHER
FT	/note= "designated J in the specification"
XX	
XX	WO200118238-A1.
PN	
XX	
XX	15-MAR-2001.
PD	
XX	
XX	11-SEP-2000; 2000WO-US24882.
XX	
PR	10-SEP-1999; 99US-0394019.
XX	
PA	(ONCO-) ONCOIMMUNIN INC.
XX	



CC detected. The present sequence is one of the peptides described in the  
exemplification of the invention.

SQ Sequence 18 AA;

Query Match 89.6%; Score 86; DB 22; Length 18;  
Best Local Similarity 94.4%; Pred. No. 3.9e-08;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDINGXPKGY 18  
| | | | | | | | | | | | | | | | | |  
Db 1 kdpwgwehdgngxpkgy 18

RESULT 4

AAG73232  
ID AAG73232 standard; Peptide; 18 AA.

AC AAG73232;

DT 14-AUG-2001 (first entry)

DE Protease binding site #166.

XX Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
KW viral infection; cancer metastasis; emphysema; arthritis;  
KW thrombosis; haemophilia.

OS Synthetic.

XX Key Location/Qualifiers  
FH Modified-site 1

FT /label= OTHER

FT /note= "modified by fluorophore"

XX WO200118238-A1.

XX 15-MAR-2001.

XX 11-SEP-2000; 2000WO-US24882.

XX 10-SEP-1999; 99US-0394019.

XX (ONCO-) ONCOIMMUNIN INC.

XX Komoriya A, Packard BS;

XX WPI; 2001-389573/41.

XX New fluorogenic compositions whose fluorescence level increases in the  
presence of active proteases, useful for detecting and localizing  
protease activity in biological samples, particularly in frozen tissue  
samples -

PS Disclosure; Page 29; 86pp; English.

XX The present invention describes fluorogenic compositions which can be  
used for the detection of protease activity. This can be useful as an  
indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
thrombosis and arthritis. The fluorogenic compositions comprise a  
peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
peptide is cleaved by a protease and the fluorophores can then be  
detected. The present sequence is one of the peptides described in the  
exemplification of the invention.

SQ Sequence 18 AA;

Query Match

Best Local Similarity 87.5%; Score 84; DB 22; Length 18;

Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDINGXPKGY 18  
| | | | | | | | | | | | | | | | | |  
Db 1 kdpwgwehdgngxpkgy 18

RESULT 5

AAG73325  
ID AAG73325 standard; Peptide; 18 AA.

AC AAG73325;

DT 14-AUG-2001 (first entry)

DE Protease indicator compound peptide #54.

XX Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
KW viral infection; cancer metastasis; emphysema; arthritis;  
KW thrombosis; haemophilia.

OS Synthetic.

XX Key Location/Qualifiers  
FH Modified-site 4

FT /label= OTHER

FT /note= "designated J in the specification"

FT 14

FT /label= OTHER

FT /note= "designated J in the specification"

XX WO200118238-A1.

XX 15-MAR-2001.

XX 11-SEP-2000; 2000WO-US24882.

XX 10-SEP-1999; 99US-0394019.

XX (ONCO-) ONCOIMMUNIN INC.

XX Komoriya A, Packard BS;

XX WPI; 2001-389573/41.

XX New fluorogenic compositions whose fluorescence level increases in the  
presence of active proteases, useful for detecting and localizing  
protease activity in biological samples, particularly in frozen tissue  
samples -

PS Claim 4; Page 71; 86pp; English.

XX The present invention describes fluorogenic compositions which can be  
used for the detection of protease activity. This can be useful as an  
indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
thrombosis and arthritis. The fluorogenic compositions comprise a  
peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
peptide is cleaved by a protease and the fluorophores can then be  
detected. The present sequence is one of the peptides described in the  
exemplification of the invention.

SQ Sequence 18 AA;

Query Match 85.4%; Score 82; DB 22; Length 18;

Best Local Similarity 94.4%; Pred. No. 1.9e-07;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDINGXPKGY 18

| | | | | | | | | | | | | | | | | |

Db 1 kdpwgwehdgngxpkgy 18

RESULT 6

AAG73231

Tue Jul 16 16:10:13 2002

thrombosis; haemophilia.  
Synthetic.  
Key Location/Qualifiers  
Modified-site 4 /label= OTHER  
/note= "designated J in the specification"  
Modified-site 14  
/label= OTHER  
/note= "designated J in the specification"  
WO200118238-A1.  
15-MAR-2001.  
11-SEP-2000; 2000WO-US24882.  
10-SEP-1999; 99US-0394019.  
(ONCO-) ONCOIMMUNIN INC.  
Komoriya A, Packard BS;  
WPI; 2001-389573/41.  
New fluorogenic compositions whose fluorescence level increases in the presence of active proteases, useful for detecting and localizing protease activity in biological samples, particularly in frozen tissue samples  
Claim 4; Page 71; 86pp; English.  
The present invention describes fluorogenic compositions which can be used for the detection of protease activity. This can be useful as an indicator of viral infection, cancer metastasis, haemophilia, emphysema, thrombosis and arthritis. The fluorogenic compositions comprise a peptide, a peptide spacer and a donor and an acceptor fluorophore. The peptide is cleaved by a protease and the fluorophores can then be detected. The present sequence is one of the peptides described in the exemplification of the invention.  
Sequence 18 AA;  
Query Match 83.3%; Score 80; DB 22; Length 18;  
Best Local Similarity 88.9%; Pred. No. 4.3e-07;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 KDPXGLEHDGNGXPKGY 18  
| | | | | | | | | | | | | | | | | |  
Db 1 kdpxyvhdgngxpkgy 18  
RESULT 8  
AAG73234  
ID AAG73234 standard; Peptide; 18 AA.  
XX AC AAG73234;  
XX DT 14-AUG-2001 (first entry)  
XX DE Protease binding site #165.  
XX DE Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
KW viral infection; cancer metastasis; emphysema; arthritis;  
KW thrombosis; haemophilia.  
XX OS Synthetic.  
XX FH Key Location/Qualifiers  
FT Modified-site 1 /label= OTHER  
FT /note= "modified by fluorophore"  
WO200118238-A1.  
15-MAR-2001.  
11-SEP-2000; 2000WO-US24882.  
10-SEP-1999; 99US-0394019.  
(ONCO-) ONCOIMMUNIN INC.  
Komoriya A, Packard BS;  
WPI; 2001-389573/41.  
New fluorogenic compositions whose fluorescence level increases in the presence of active proteases, useful for detecting and localizing protease activity in biological samples, particularly in frozen tissue samples  
Disclosure; Page 29; 86pp; English.  
The present invention describes fluorogenic compositions which can be used for the detection of protease activity. This can be useful as an indicator of viral infection, cancer metastasis, haemophilia, emphysema, thrombosis and arthritis. The fluorogenic compositions comprise a peptide, a peptide spacer and a donor and an acceptor fluorophore. The peptide is cleaved by a protease and the fluorophores can then be detected. The present sequence is one of the peptides described in the exemplification of the invention.  
Sequence 18 AA;  
Query Match 83.3%; Score 80; DB 22; Length 18;  
Best Local Similarity 83.3%; Pred. No. 4.3e-07;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 KDPXGLEHDGNGXPKGY 18  
| | | | | | | | | | | | | | | | | |  
Db 1 kdpcegtgdgngcpgky 18  
RESULT 7  
AAG73328  
ID AAG73328 standard; Peptide; 18 AA.  
XX AC AAG73328;  
XX DT 14-AUG-2001 (first entry)  
XX DE Protease indicator compound peptide #57.  
XX DE Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
KW viral infection; cancer metastasis; emphysema; arthritis;  
KW

thrombosis; haemophilia.  
Synthetic.  
Key Location/Qualifiers  
Modified-site 4 /label= OTHER  
/note= "designated J in the specification"  
Modified-site 14  
/label= OTHER  
/note= "designated J in the specification"  
WO200118238-A1.  
15-MAR-2001.  
11-SEP-2000; 2000WO-US24882.  
10-SEP-1999; 99US-0394019.  
(ONCO-) ONCOIMMUNIN INC.  
Komoriya A, Packard BS;  
WPI; 2001-389573/41.  
New fluorogenic compositions whose fluorescence level increases in the presence of active proteases, useful for detecting and localizing protease activity in biological samples, particularly in frozen tissue samples  
Claim 4; Page 71; 86pp; English.  
The present invention describes fluorogenic compositions which can be used for the detection of protease activity. This can be useful as an indicator of viral infection, cancer metastasis, haemophilia, emphysema, thrombosis and arthritis. The fluorogenic compositions comprise a peptide, a peptide spacer and a donor and an acceptor fluorophore. The peptide is cleaved by a protease and the fluorophores can then be detected. The present sequence is one of the peptides described in the exemplification of the invention.  
Sequence 18 AA;  
Query Match 83.3%; Score 80; DB 22; Length 18;  
Best Local Similarity 88.9%; Pred. No. 4.3e-07;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 KDPXGLEHDGNGXPKGY 18  
| | | | | | | | | | | | | | | | | |  
Db 1 kdpxyvhdgngxpkgy 18  
RESULT 8  
AAG73234  
ID AAG73234 standard; Peptide; 18 AA.  
XX AC AAG73234;  
XX DT 14-AUG-2001 (first entry)  
XX DE Protease binding site #168.  
XX DE Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
KW viral infection; cancer metastasis; emphysema; arthritis;  
KW thrombosis; haemophilia.  
XX OS Synthetic.  
XX FH Key Location/Qualifiers  
FT Modified-site 1 /label= OTHER  
FT

FT XX /note= "modified by fluorophore"

PN WO200118238-A1.

XX 15-MAR-2001.

XX 11-SEP-2000; 2000WO-US24882.

PF PD 10-SEP-1999; 99US-0394019.

PR PA (ONCO-) ONCOIMMUNIN INC.

XX PI Komoriya A, Packard BS;

XX DR WPI; 2001-389573/41.

XX PT New fluorogenic compositions whose fluorescence level increases in the presence of active proteases, useful for detecting and localizing protease activity in biological samples, particularly in frozen tissue samples.

PS Disclosure; Page 29; 86pp; English.

CC The present invention describes fluorogenic compositions which can be used for the detection of protease activity. This can be useful as an indicator of viral infection, cancer metastasis, haemophilia, emphysema, thrombosis and arthritis. The fluorogenic compositions comprise a peptide, a peptide spacer and a donor and an acceptor fluorophore. The peptide is cleaved by a protease and the fluorophores can then be detected. The present sequence is one of the peptides described in the exemplification of the invention.

XX SQ Sequence 18 AA;

Query Match 81.2%; Score 78; DB 22; Length 18;  
Best Local Similarity 77.8%; Pred. No. 9.5e-07;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDINGXPKGY 18  
| | | | | | | | | | | | | | | | | |

Db 1 kdpccgyvhdgngcpkgy 18

RESULT 9

AAW82112

ID AAW82112 standard; peptide; 18 AA.

XX AC AAW82112;

XX DT 18-FEB-1999 (first entry)

XX DE Fluorogenic protease indicator CPP32 substrate #1.

XX KW Protease activity; fluorophore; detection; fluorogenic; cellular uptake; conformation change.

XX OS Synthetic.

FT Key Location/Qualifiers

FT Modified-site 1

FT /note= "Lys residue modified by presence of Fmoc (9-fluorenylmethoxycarbonyl)"

FT Modified-site 4

FT /note= "epsilon-aminocaproic acid, labelled as amino acid J in the specification"

FT Modified-site 14

FT /note= "epsilon-aminocaproic acid, labelled as amino acid J in the specification"

XX WO9837226-A1.

XX PN 27-AUG-1998.

XX 20-FEB-1998; 98WO-US03000.

XX 20-FEB-1997; 97US-0802981.

PR XX (ONCO-) ONCOIMMUNIN INC.

XX PI Komoriya A, Packard BS;

XX DR WPI; 1998-467579/40.

XX PT New fluorogenic compositions - containing 2 fluorophores separated by a peptide comprising a protease binding site, used for detecting protease activity in samples.

PS Disclosure; Page 24; 90pp; English.

XX CC AAW82023-W82240 are peptides used in the construction of a fluorogenic composition which is used for the detection of protease activity in biological samples. The products can be used for the detection of conformation changes in nucleic acids, oligosaccharides, polysaccharides, proteins, peptides, lipids, phospholipids, glycoproteins, steroids or polymers. In addition, attachment of a hydrophobic group to a molecule can be used to enhance uptake by cells. The composition is composed of P = peptide comprising a protease binding site for the protease, F1, F2 peptides = fluorophores where F1 is attached to the amino terminal amino acid and F2 is attached to the carboxyl terminal amino acid and S1, S2 peptides = when present, are peptide spacers where S1, when present, is attached to the amino terminal acid, and S2, when present, is attached to the carboxyl terminal amino acid.

XX SQ Sequence 18 AA;

Query Match 76.0%; Score 73; DB 19; Length 18;  
Best Local Similarity 88.9%; Pred. No. 7e-06;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDINGXPKGY 18  
| | | | | | | | | | | | | | | | | |

Db 1 kdpdxgdevdngxpkgy 18

RESULT 10

AAW823300

ID AAG73300 standard; Peptide; 18 AA.

XX AC AAG73300;

XX DT 14-AUG-2001 (first entry)

XX DE Protease indicator compound peptide #29.

XX KW Protease detection; peptide cleavage; enzyme activity; fluorogenic; viral infection; cancer metastasis; emphysema; arthritis; thrombosis; haemophilia.

XX OS Synthetic.

FT Key Location/Qualifiers

FT Modified-site 1

FT /label= OTHER

FT /note= "modified by Fa"

FT Modified-site 4

FT /label= OTHER

FT Modified-site 14

FT /note= "designated J in the specification"

FT /label= OTHER

FT /note= "designated J in the specification"

XX WO200118238-A1.

[illegible]

PT Presence of active proteases, useful for detecting and localizing  
PT protease activity in biological samples, particularly in frozen tissue  
PT samples -

PS Disclosure; Page 23; 86pp; English.

CC The present invention describes fluorogenic compositions which can be  
CC used for the detection of protease activity. This can be useful as an  
CC indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
CC thrombosis and arthritis. The fluorogenic compositions comprise a  
CC peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
CC peptide is cleaved by a protease and the fluorophores can then be  
CC detected. The present sequence is one of the peptides described in the  
CC exemplification of the invention.

XX Sequence 18 AA;

Query Match 74.0%; Score 71; DB 22; Length 18;  
Best Local Similarity 77.8%; Pred. No. 1.5e-05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 KDPXGLEHDGNGXPKGY 18  
||| | | |||| | ||||  
DB 1 kdpcgdevdgingcpgky 18

RESULT 13

AAG73086  
ID AAG73086 standard; Peptide; 18 AA.

AC AAG73086;

DT 14-AUG-2001 (first entry)

DE Protease binding site #20.

XX Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
XX viral infection; cancer metastasis; emphysema; arthritis;  
KW thrombosis; haemophilia.

OS Synthetic.

XX Key Location/Qualifiers  
FH Modified-site 1  
FT /label= OTHER  
FT /note= "modified by fluorophore"

XX WO200118238-A1.

XX 15-MAR-2001.

XX 11-SEP-2000; 2000WO-US24882.

XX 10-SEP-1999; 99US-0394019.

XX (ONCO-) ONCOIMMUNIN INC.

XX Komoriya A, Packard BS;

XX WPI; 2001-389573/41.

PT New fluorogenic compositions whose fluorescence level increases in the  
PT presence of active proteases, useful for detecting and localizing  
PT protease activity in biological samples, particularly in frozen tissue  
PT samples -

PS Disclosure; Page 23; 86pp; English.

CC The present invention describes fluorogenic compositions which can be  
CC used for the detection of protease activity. This can be useful as an  
CC indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
CC thrombosis and arthritis. The fluorogenic compositions comprise a

CC peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
CC peptide is cleaved by a protease and the fluorophores can then be  
CC detected. The present sequence is one of the peptides described in the  
CC exemplification of the invention.

XX Sequence 18 AA;

Query Match 74.0%; Score 71; DB 22; Length 18;  
Best Local Similarity 77.8%; Pred. No. 1.5e-05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 KDPXGLEHDGNGXPKGY 18  
||| | | |||| | ||||  
DB 1 kdpcgdevdgingcpgky 18

RESULT 14

AAG73090

ID AAG73090 standard; Peptide; 18 AA.

AC AAG73090;

DT 14-AUG-2001 (first entry)

DE Protease binding site #24.

XX Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
XX viral infection; cancer metastasis; emphysema; arthritis;  
KW thrombosis; haemophilia.

OS Synthetic.

XX Key Location/Qualifiers

FH Modified-site 1

FT /label= OTHER

FT /note= "modified by fluorophore"

FT Misc-difference 6

FT /note= "D-form residue"

FT Misc-difference 9

FT /note= "D-form residue"

XX WO200118238-A1.

XX 15-MAR-2001.

XX 11-SEP-2000; 2000WO-US24882.

XX 10-SEP-1999; 99US-0394019.

XX (ONCO-) ONCOIMMUNIN INC.

XX Komoriya A, Packard BS;

XX WPI; 2001-389573/41.

PT New fluorogenic compositions whose fluorescence level increases in the  
PT presence of active proteases, useful for detecting and localizing  
PT protease activity in biological samples, particularly in frozen tissue  
PT samples -

PS Disclosure; Page 23; 86pp; English.

CC The present invention describes fluorogenic compositions which can be  
CC used for the detection of protease activity. This can be useful as an  
CC indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
CC thrombosis and arthritis. The fluorogenic compositions comprise a  
CC peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
CC peptide is cleaved by a protease and the fluorophores can then be  
CC detected. The present sequence is one of the peptides described in the  
CC exemplification of the invention.

XX Sequence 18 AA;

Tue Jul 16 16:10:13 2002

Search completed: July 16, 2002, 11:21:27  
Job time: 365 sec

Query Match 74.0%; Score 71; DB 22; Length 18;  
Best Local Similarity 77.8%; Pred. No. 1.5e-05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDGNGXPKGY 18  
||| | | |||| |  
DB 1 kdpdgdevdgingcpkgy 18

RESULT 15

AAG73091  
ID AAG73091 standard; Peptide; 18 AA.

XX  
AC AAG73091;

DT 14-AUG-2001 (first entry)

XX Protease binding site #25.

XX Protease detection; peptide cleavage; enzyme activity; fluorogenic;  
KW viral infection; cancer metastasis; emphysema; arthritis;  
KW thrombosis; haemophilia.

XX Synthetic.

XX Key Location/Qualifiers  
FH Modified-site 1  
FT /label= OTHER  
FT /note= "modified by fluorophore"

XX WO200118238-A1.

XX PD 15-MAR-2001.

XX 11-SEP-2000; 2000WO-US24882.

XX PR 10-SEP-1999; 99US-0394019.

XX PA (ONCO-) ONCOIMMUNIN INC.

XX Komoriya A, Packard BS;

XX WPI; 2001-389573/41.

XX New fluorogenic compositions whose fluorescence level increases in the  
PT presence of active proteases, useful for detecting and localizing  
PT protease activity in biological samples, particularly in frozen tissue  
PT samples -

XX Disclosure; Page 23; 86pp; English.

XX The present invention describes fluorogenic compositions which can be  
CC used for the detection of protease activity. This can be useful as an  
CC indicator of viral infection, cancer metastasis, haemophilia, emphysema,  
CC thrombosis and arthritis. The fluorogenic compositions comprise a  
CC peptide, a peptide spacer and a donor and an acceptor fluorophore. The  
CC peptide is cleaved by a protease and the fluorophores can then be  
CC detected. The present sequence is one of the peptides described in the  
CC exemplification of the invention.

XX Sequence 18 AA;

Query Match 74.0%; Score 71; DB 22; Length 18;  
Best Local Similarity 77.8%; Pred. No. 1.5e-05;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDGNGXPKGY 18  
||| | | |||| |  
DB 1 kdpdgdevdgingcpkgy 18



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:

July 16, 2002, 11:15:22 ; Search time 12.98 seconds  
(without alignments)  
33.872 Million cell updates/sec

Title: US-09-394-019A-248

Perfect score: 96

Sequence: 1 KDPXGLEHGDINGXPKGY 18

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	65	67.7	18	3	US-08-802-981-55		Sequence 55, Appl
2	64	66.7	18	3	US-08-802-981-54		Sequence 54, Appl
3	61	63.5	18	3	US-08-802-981-69		Sequence 69, Appl
4	60	62.5	18	3	US-08-802-981-53		Sequence 53, Appl
5	60	62.5	18	3	US-08-802-981-57		Sequence 57, Appl
6	59	61.5	18	3	US-08-802-981-56		Sequence 56, Appl
7	56	58.3	18	3	US-08-802-981-66		Sequence 66, Appl
8	56	58.3	18	3	US-08-802-981-68		Sequence 68, Appl
9	54	56.2	18	3	US-08-802-981-49		Sequence 49, Appl
10	54	56.2	18	3	US-08-802-981-50		Sequence 50, Appl
11	51	53.1	18	3	US-08-802-981-67		Sequence 67, Appl
12	50.5	52.6	19	3	US-08-802-981-59		Sequence 59, Appl
13	48	50.0	18	3	US-08-802-981-173		Sequence 173, Appl
14	48	50.0	18	3	US-08-802-981-174		Sequence 174, Appl
15	47.5	49.5	19	3	US-08-802-981-76		Sequence 76, Appl
16	44.5	46.4	19	3	US-08-802-981-58		Sequence 58, Appl
17	44	45.8	19	3	US-08-802-981-52		Sequence 52, Appl
18	43.5	45.3	16	3	US-08-802-981-64		Sequence 64, Appl
19	43.5	45.3	655	4	US-09-245-808-1		Sequence 1, Appl
20	41.5	43.2	16	3	US-08-802-981-48		Sequence 48, Appl
21	40	41.7	18	3	US-08-802-981-73		Sequence 73, Appl
22	40	41.7	18	3	US-08-802-981-80		Sequence 80, Appl
23	39	40.6	19	3	US-08-802-981-72		Sequence 72, Appl
24	39	40.6	232	4	US-08-979-608A-3		Sequence 3, Appl
25	39	40.6	252	4	US-08-979-608A-4		Sequence 4, Appl
26	39	40.6	317	4	US-08-979-608A-2		Sequence 2, Appl
27	38.5	40.1	16	3	US-08-802-981-63		Sequence 63, Appl

28	38.5	40.1	16	3	US-08-802-981-65	Sequence 65, Appl
29	38	39.6	16	3	US-08-802-981-71	Sequence 71, Appl
30	38	39.6	19	3	US-08-802-981-86	Sequence 86, Appl
31	38	39.6	1297	4	US-09-540-245A-17	Sequence 17, Appl
32	38	39.6	1461	2	US-08-993-228-10	Sequence 10, Appl
33	37.5	39.1	383	4	US-09-413-814-12	Sequence 12, Appl
34	37	38.5	229	3	US-09-040-285A-7	Sequence 7, Appl
35	37	38.5	2556	1	US-08-185-432-17	Sequence 17, Appl
36	37	38.5	2556	3	US-08-083-590A-20	Sequence 20, Appl
37	37	38.5	2556	3	US-08-532-981-20	Sequence 20, Appl
38	36.5	38.0	17	3	US-08-802-981-78	Sequence 78, Appl
39	36	37.5	14	3	US-08-802-981-61	Sequence 61, Appl
40	36	37.5	43	2	US-08-488-161-38	Sequence 38, Appl
41	36	37.5	43	3	US-09-273-685-38	Sequence 38, Appl
42	36	37.5	43	5	PCT-US95-11934-38	Sequence 38, Appl
43	36	37.5	185	4	US-08-975-762-11	Sequence 11, Appl
44	36	37.5	185	4	US-08-821-324-11	Sequence 11, Appl
45	36	37.5	185	4	US-09-295-028-11	Sequence 11, Appl

## ALIGNMENTS

RESULT 1  
US-08-802-981-55  
; Sequence 55, Application US/08802981  
; Patent No. 6037137  
; GENERAL INFORMATION:  
; APPLICANT: Komoriya, Akira  
; APPLICANT: Packard, Beverly S.  
; TITLE OF INVENTION: Compositions for the Detection of Enzyme  
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
; NUMBER OF SEQUENCES: 231  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/802,981  
; FILING DATE: 20-FEB-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hunter, Tom  
; REGISTRATION NUMBER: 38,498  
; REFERENCE/DOCKET NUMBER: 016865-00030005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 55:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 3  
; OTHER INFORMATION: /product= "Aib"  
; NAME/KEY: Modified-site  
; LOCATION: 4  
; OTHER INFORMATION: /product= "Acp"  
; FEATURE:  
; NAME/KEY: Modified-site

us-09-394-019a-248.ra

Tue Jul 16 16:10:13 2002

QY 1 KDPXGLEHDXGXPXKY 18  
||| | |||||  
Db 1 KDXGDEVGDXGXPXKY 18

RESULT 3  
US-08-802-981-69  
; Sequence 69, Application US/08802981  
; Patent No. 6037137  
; GENERAL INFORMATION:  
; APPLICANT: Komoriya, Akira  
; APPLICANT: Packard, Beverly S.  
; TITLE OF INVENTION: Compositions for the Detection of Enzyme  
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
; NUMBER OF SEQUENCES: 231  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/802,981  
; FILING DATE: 20-FEB-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hunter, Tom  
; REGISTRATION NUMBER: 38,498  
; REFERENCE/DOCKET NUMBER: 016865-00030005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 69:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 3  
; OTHER INFORMATION: /product= "Aib"  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 4  
; OTHER INFORMATION: /product= "Acp"  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 14  
; OTHER INFORMATION: /product= "Acp"  
; US-08-802-981-69

Query Match 63.5%; Score 61; DB 3; Length 18;  
Best Local Similarity 77.8%; Pred. No. 0.0003;  
Matches 14; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 KDPXGLEHDXGXPXKY 18  
||| | |||||  
Db 1 KDXGDEVGDXGXPXKY 18

RESULT 4  
US-08-802-981-53  
; Sequence 53, Application US/08802981  
; Patent No. 6037137

LOCATION: 14  
OTHER INFORMATION: /product= "Acp"  
US-08-802-981-55

Query Match 67.7%; Score 65; DB 3; Length 18;  
Best Local Similarity 83.3%; Pred. No. 6.2e-05;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KDPXGLEHDXGXPXKY 18  
||| | |||||  
Db 1 KDXGDEVGDXGXPXKY 18

RESULT 2  
US-08-802-981-54  
; Sequence 54, Application US/08802981  
; Patent No. 6037137  
; GENERAL INFORMATION:  
; APPLICANT: Komoriya, Akira  
; APPLICANT: Packard, Beverly S.  
; TITLE OF INVENTION: Compositions for the Detection of Enzyme  
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
; NUMBER OF SEQUENCES: 231  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/802,981  
; FILING DATE: 20-FEB-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hunter, Tom  
; REGISTRATION NUMBER: 38,498  
; REFERENCE/DOCKET NUMBER: 016865-00030005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 54:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 3  
; OTHER INFORMATION: /product= "Aib"  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 4  
; OTHER INFORMATION: /product= "Acp"  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 14  
; OTHER INFORMATION: /product= "Acp"  
; US-08-802-981-54

Query Match 66.7%; Score 64; DB 3; Length 18;  
Best Local Similarity 83.3%; Pred. No. 9.2e-05;  
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

;
; GENERAL INFORMATION:
; APPLICANT: Komoriya, Akira
; TITLE OF INVENTION: Compositions for the Detection of Enzyme
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof
; NUMBER OF SEQUENCES: 231
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,981
; FILING DATE: 20-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 016865-000300US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /product= "Alb"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 4
; OTHER INFORMATION: /product= "Acp"
; NAME/KEY: Modified-site
; LOCATION: 14
; OTHER INFORMATION: /product= "Acp"
;
; US-08-802-981-53
;
; Query Match 62.5%; Score 60; DB 3; Length 18;
; Best Local Similarity 77.8%; Pred. No. 0.00044;
; Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
;
; QY 1 KDPXGLEHDGNGXPXPKGY 18
; Db 1 KDXGNEVDGIDGXPXPKGY 18
;
; RESULT 5
; US-08-802-981-57
; Sequence 57, Application US/08802981
; Patent No. 6037137
; GENERAL INFORMATION:
; APPLICANT: Komoriya, Akira
; TITLE OF INVENTION: Compositions for the Detection of Enzyme
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof
; NUMBER OF SEQUENCES: 231
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```

;
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,981
; FILING DATE: 20-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 016865-000300US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /product= "Alb"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 4
; OTHER INFORMATION: /product= "Acp"
; NAME/KEY: Modified-site
; LOCATION: 14
; OTHER INFORMATION: /product= "Acp"
;
; US-08-802-981-57
;
; Query Match 62.5%; Score 60; DB 3; Length 18;
; Best Local Similarity 77.8%; Pred. No. 0.00044;
; Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
;
; QY 1 KDPXGLEHDGNGXPXPKGY 18
; Db 1 KDXGNEVDGIDGXPXPKGY 18
;
; RESULT 6
; US-08-802-981-56
; Sequence 56, Application US/08802981
; Patent No. 6037137
; GENERAL INFORMATION:
; APPLICANT: Komoriya, Akira
; TITLE OF INVENTION: Compositions for the Detection of Enzyme
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof
; NUMBER OF SEQUENCES: 231
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

us-09-394-019a-248.ra1

Tue Jul 16 16:10:13 2002

TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 3  
OTHER INFORMATION: /product= "Aib"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 4  
OTHER INFORMATION: /product= "Acp"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 14  
OTHER INFORMATION: /product= "Acp"  
US-08-802-981-66

Query Match 58.3%; Score 56; DB 3; Length 18;  
Best Local Similarity 72.2%; Pred. No. 0.0021;  
Matches 13; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDINGXPKGY 18  
Db 1 KDXXGVVADGIDGXPXGY 18

RESULT 8  
US-08-802-981-68  
Sequence 68, Application US/08802981  
Patent No. 6037137  
GENERAL INFORMATION:  
APPLICANT: Komoriya, Akira  
APPLICANT: Packard, Beverly S.  
TITLE OF INVENTION: Compositions for the Detection of Enzyme  
TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
NUMBER OF SEQUENCES: 231  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,981  
FILING DATE: 20-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 016865-0003000US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,981  
FILING DATE: 20-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 016865-0003000US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 56:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 3  
OTHER INFORMATION: /product= "Aib"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 4  
OTHER INFORMATION: /product= "Acp"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 14  
OTHER INFORMATION: /product= "Acp"  
US-08-802-981-56

Query Match 61.5%; Score 59; DB 3; Length 18;  
Best Local Similarity 77.8%; Pred. No. 0.00065;  
Matches 14; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDINGXPKGY 18  
Db 1 KDXXGVVADGIDGXPXGY 18

RESULT 7  
US-08-802-981-66  
Sequence 66, Application US/08802981  
Patent No. 6037137  
GENERAL INFORMATION:  
APPLICANT: Komoriya, Akira  
APPLICANT: Packard, Beverly S.  
TITLE OF INVENTION: Compositions for the Detection of Enzyme  
TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
NUMBER OF SEQUENCES: 231  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,981  
FILING DATE: 20-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 016865-0003000US  
TELECOMMUNICATION INFORMATION:

FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 3  
OTHER INFORMATION: /product= "Aib"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 4  
OTHER INFORMATION: /product= "Acp"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 14  
OTHER INFORMATION: /product= "Acp"  
US-08-802-981-68

Query Match 58.38; Score 56; DB 3; Length 18;  
Best Local Similarity 72.28; Pred. No. 0.0021;  
Matches 13; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KDPXGLEHGDINGXPKGY 18  
||| :|||||  
Db 1 KDXGYVANGINGXPKGY 18

RESULT 9  
US-08-802-981-49  
Sequence 49, Application US/08802981  
Patent No. 6037137  
GENERAL INFORMATION:  
APPLICANT: Komoriya, Akira  
TITLE OF INVENTION: Compositions for the Detection of Enzyme  
TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
NUMBER OF SEQUENCES: 231  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,981  
FILING DATE: 20-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 016865-000300US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /product= "OTHER"  
OTHER INFORMATION: /note= "Xaa = 9-fluorenylmethoxycarbonyl (Fmoc) group attached to the alpha-amino group of Lys"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 3

OTHER INFORMATION: /product= "Aib"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 4  
OTHER INFORMATION: /product= "Acp"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 14  
OTHER INFORMATION: /product= "Acp"  
US-08-802-981-49

Query Match 56.28; Score 54; DB 3; Length 18;  
Best Local Similarity 76.58; Pred. No. 0.0046;  
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 DPXGLEHGDINGXPKGY 18  
||| :|||||  
Db 2 DXXGDEVGDIDGXPKGY 18

RESULT 10  
US-08-802-981-50  
Sequence 50, Application US/08802981  
Patent No. 6037137  
GENERAL INFORMATION:  
APPLICANT: Komoriya, Akira  
TITLE OF INVENTION: Compositions for the Detection of Enzyme  
TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
NUMBER OF SEQUENCES: 231  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/802,981  
FILING DATE: 20-FEB-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 016865-000300US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /product= "OTHER"  
OTHER INFORMATION: /note= "Xaa = benzylloxycarbonyl (Z) group attached to the  
OTHER INFORMATION: group of Lys"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 3  
OTHER INFORMATION: /product= "Aib"  
FEATURE:  
NAME/KEY: Modified-site

us-09-394-019a-248.ra1

Tue Jul 16 16:10:13 2002

```

;
; LOCATION: 4
; OTHER INFORMATION: /product= "Acp"
;
; NAME/KEY: Modified-site
; LOCATION: 14
; OTHER INFORMATION: /product= "Acp"
;
US-08-802-981-50

Query Match 53.1%; Score 51; DB 3; Length 18;
Best Local Similarity 66.7%; Pred. No. 0.015;
Matches 12; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 KDPXGLEHGDINGXPKGY 18
   ||| :||:|||||
Db 1 KDXXGYVANGIDGXPKGY 18

RESULT 12
US-08-802-981-59
; Sequence 59, Application US/08802981
; Patent No. 6037137
; GENERAL INFORMATION:
; APPLICANT: Komoriya, Akira
; APPLICANT: Packard, Beverly S.
; TITLE OF INVENTION: Compositions for the Detection of Enzyme
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof
; NUMBER OF SEQUENCES: 231
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,981
; FILING DATE: 20-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 016865-0003000US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /product= "Aib"
;
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 4
; OTHER INFORMATION: /product= "Acp"
;
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /product= "Acp"
;
US-08-802-981-59

```

```

Query Match 52.6%; Score 50.5; DB 3; Length 19;
Best Local Similarity 73.7%; Pred. No. 0.019;
Matches 14; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 KDPXGL-EHDGNGXPKGY 18
   ||| :||:|||||
Db 1 KDXXGWDEYDGDGXPKGY 19

```

```

;
; LOCATION: 4
; OTHER INFORMATION: /product= "Acp"
;
; NAME/KEY: Modified-site
; LOCATION: 14
; OTHER INFORMATION: /product= "Acp"
;
US-08-802-981-50

Query Match 56.2%; Score 54; DB 3; Length 18;
Best Local Similarity 76.5%; Pred. No. 0.0046;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 DPXGLEHGDINGXPKGY 18
   ||| :||:|||||
Db 2 DXGDEYDGDGXPKGY 18

RESULT 11
US-08-802-981-67
; Sequence 67, Application US/08802981
; Patent No. 6037137
; GENERAL INFORMATION:
; APPLICANT: Komoriya, Akira
; APPLICANT: Packard, Beverly S.
; TITLE OF INVENTION: Compositions for the Detection of Enzyme
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof
; NUMBER OF SEQUENCES: 231
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,981
; FILING DATE: 20-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 016865-0003000US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /product= "Aib"
;
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 4
; OTHER INFORMATION: /product= "Acp"
;
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 14
; OTHER INFORMATION: /product= "Acp"
;
US-08-802-981-67

```

RESULT 13  
US-08-802-981-173  
; Sequence 173, Application US/08802981  
; Patent No. 6037137  
; GENERAL INFORMATION:  
; APPLICANT: Komoriya, Akira  
; APPLICANT: Packard, Beverly S.  
; TITLE OF INVENTION: Compositions for the Detection of Enzyme  
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
; NUMBER OF SEQUENCES: 231  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION NUMBER: US/08/802,981  
; FILING DATE: 20-FEB-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hunter, Tom  
; REGISTRATION NUMBER: 38,498  
; REFERENCE/DOCKET NUMBER: 016865-000300US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 173:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 1  
; OTHER INFORMATION: /product= "OTHER"  
; OTHER INFORMATION: /note= "Xaa = 9-fluorenylmethoxycarbonyl (Fmoc) group attached  
; OTHER INFORMATION: to the alpha-amino group and 5'-carboxytetramethylrhodamine  
; OTHER INFORMATION: attached to the epsilon-amino group of Lys"  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 3  
; OTHER INFORMATION: /product= "Aib"  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 4  
; OTHER INFORMATION: /product= "Acp"  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 14  
; OTHER INFORMATION: /product= "Acp"  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 16  
; OTHER INFORMATION: /product= "OTHER"  
; OTHER INFORMATION: /note= "Xaa = 5'-carboxytetramethylrhodamine attached to the  
; OTHER INFORMATION: epsilon-amino group of Lys"  
US-08-802-981-173

Query Match 50.0%; Score 48; DB 3; Length 18;  
Best Local Similarity 70.6%; Pred. No. 0.047;  
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 DPXGLEHGDINGXPXY 18

Db 2 DXSGDEVGDIDGXPXY 18

RESULT 14  
US-08-802-981-174  
; Sequence 174, Application US/08802981  
; Patent No. 6037137  
; GENERAL INFORMATION:  
; APPLICANT: Komoriya, Akira  
; APPLICANT: Packard, Beverly S.  
; TITLE OF INVENTION: Compositions for the Detection of Enzyme  
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof  
; NUMBER OF SEQUENCES: 231  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION NUMBER: US/08/802,981  
; FILING DATE: 20-FEB-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hunter, Tom  
; REGISTRATION NUMBER: 38,498  
; REFERENCE/DOCKET NUMBER: 016865-000300US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 174:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 1  
; OTHER INFORMATION: /product= "OTHER"  
; OTHER INFORMATION: /note= "Xaa = benzyloxycarbonyl (Z) group attached to the  
; OTHER INFORMATION: alpha-amino group and 5'-carboxytetramethylrhodamine attac  
; OTHER INFORMATION: the epsilon-amino group of Lys"  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 3  
; OTHER INFORMATION: /product= "Aib"  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 4  
; OTHER INFORMATION: /product= "Acp"  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 14  
; OTHER INFORMATION: /product= "Acp"  
; FEATURE:  
; NAME/KEY: Modified-site  
; LOCATION: 16  
; OTHER INFORMATION: /product= "OTHER"  
; OTHER INFORMATION: /note= "Xaa = 5'-carboxytetramethylrhodamine attached to t  
; OTHER INFORMATION: epsilon-amino group of Lys"  
US-08-802-981-174

Query Match 50.0%; Score 48; DB 3; Length 18;

us-09-394-019a-248.rai

Tue Jul 16 16:10:13 2002

```

Best Local Similarity 70.6%; Pred. No. 0.047;
Matches 12; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 DPXGLEHGXPKGY 18
   ||| |||:|||||
Db 2 DXGDEVGDGXPKGY 18

RESULT 15
US-08-802-981-76
; Sequence 76, Application US/08802981
; Patent No. 6037137
; GENERAL INFORMATION:
; APPLICANT: Komoriya, Akira
; APPLICANT: Packard, Beverly S.
; TITLE OF INVENTION: Compositions for the Detection of Enzyme
; TITLE OF INVENTION: Activity in Biological Samples and Methods of Use Thereof
; NUMBER OF SEQUENCES: 231
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,981
; FILING DATE: 20-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 016865-000300US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /product= "Aib"
;
US-08-802-981-76

```

```

Query Match 49.5%; Score 47.5; DB 3; Length 19;
Best Local Similarity 52.6%; Pred. No. 0.061;
Matches 10; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

Qy 1 KDPXGLEHD-GINGXPKGY 18
   ||| |||:|||||
Db 1 KDXGGIETDSGVDDGPKGY 19

```

Search completed: July 16, 2002, 11:20:29  
Job time: 307 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2002, 11:15:22 ; Search time 14.62 Seconds  
(without alignments)  
118.304 Million cell updates/sec

Title: US-09-394-019A-248

Perfect score: 96  
Sequence: 1 KDPXGLEHNGXPKGY 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	45	46.9	135	2	G87653		hypothetical prote
2	44.5	46.4	542	2	E72212		formate--tetrahydr
3	43.5	45.3	1146	2	T09112		probable sensor ki
4	43.5	45.3	1146	2	A82174		sensory box sensor
5	43	44.8	289	2	A43562		homeotic protein H
6	43	44.8	333	2	S42424		hypothetical prote
7	42.5	44.3	370	2	D83818		heat-shock protein
8	42	43.8	73	2	I47089		keratin type II -
9	42	43.8	99	2	A41605		homeotic protein H
10	42	43.8	118	2	JC2568		myr protein - Rhl
11	42	43.8	129	2	I56195		gene Tap-1 protein
12	42	43.8	196	2	AH2441		hypothetical prote
13	42	43.8	200	2	H82320		HAMI protein VC045
14	42	43.8	427	2	I51580		XFRK2 protein - Af
15	42	43.8	497	1	WMBELM		membrane protein L
16	42	43.8	577	2	A37779		histocompatibility
17	42	43.8	725	2	S13426		multidrug resistan
18	41.5	43.2	327	2	A10522		pdxA-like protein
19	41	42.7	94	2	F82636		hypothetical prote
20	41	42.7	327	2	G83921		hypothetical prote
21	41	42.7	448	2	A36311		70K UI small nucle
22	41	42.7	519	2	S78850		hypothetical prote
23	41	42.7	798	2	S29815		N-ras upstream pro
24	41	42.7	798	2	S11210		probable unr prote
25	41	42.7	902	2	A55543		cmaA protein - Pse
26	41	42.7	1146	2	B35962		protein-tyrosine k
27	41	42.7	1182	2	A35962		protein-tyrosine k
28	40.5	42.2	1451	2	D64203		DNA polymerase III
29	40	41.7	185	2	T12772		conserved hypothet

ALIGNMENTS

RESULT 1

G87653

hypothetical protein CC3265 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: G87653

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, N.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: G87653

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-135 <STO>

A:Cross-references: GB:AE005673; NID:gl3424955; PIDN:AAK25227.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC3265

Query Match 46.9%; Score 45; DB 2; Length 135;  
Best Local Similarity 50.0%; Pred. No. 2.4;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 5 GLEHNGXPKGY 18

Db 89 GMDFDGTGLPKGW 102

1:: 11 1 111:

RESULT 2

E72212

formate--tetrahydrofolate ligase - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: E72212

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; M. Garrett, M.N.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MUID:99287316

A:Accession: E72212

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-542 <ARN>

A:Cross-references: GB:AE001815; GB:AE000512; NID:g4982341; PIDN:AAD36830.1; PID:g498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1766

C:Superfamily: formate--tetrahydrofolate ligase; formate--tetrahydrofolate ligase hom

```

A43562
homeotic protein Hox D8 - mouse
N;Alternate names: homeotic protein Hox 4.3
C;Species: Mus musculus (house mouse)
C;Date: 11-Dec-1992 #sequence_revision 11-Dec-1992 #text_change 30-May-1997
C;Accession: A43562
R;Izpisua-Belmonte, J.C.; Dolle, P.; Renucci, A.; Zappavigna, V.; Falkenstein, H.; Du
Development 110, 733-745, 1990
A;Title: Primary structure and embryonic expression pattern of the mouse Hox-4.3 home
A;Reference number: A43562; MUID:91209232
A;Accession: A43562
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-289 <IZP>
A;Superfamily: homeotic protein Hox A7; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;196-252/Domain: homeobox homology <HOX>

Query Match 44.8%; Score 43; DB 2; Length 289;
Best Local Similarity 53.3%; Pred. No. 12;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 KDPXGLEHDXGPKG 15
DB 273 KEASGLEEDGAECCP 287
I: |||||
: |||||

RESULT 6
S42424
hypothetical protein Y - Mycobacterium smegmatis
C;Species: Mycobacterium smegmatis
C;Date: 07-May-1998 #sequence_revision 15-May-1998 #text_change 22-Oct-1999
C;Accession: S42424; S31804
R;Cirillo, J.D.; Weisbrod, T.R.; Pascopella, L.; Bloom, B.R.; Jacobs Jr., W.R.
Mol. Microbiol. 11, 629-639, 1994
A;Title: Isolation and characterization of the aspartokinase and aspartate semialdehy
A;Reference number: S42421; MUID:94254720
A;Accession: S42424
A;Molecule type: DNA
A;Residues: 1-333 <CIR>
A;Cross-references: EMBL:Z17372; NID:644506; PIDN:CAA78987.1; PID:9581353
A;Note: the authors translated the initiation codon GTG for residue 1 as Val
C;Genetics:
A;Start codon: GTG

Query Match 44.8%; Score 43; DB 2; Length 333;
Best Local Similarity 40.0%; Pred. No. 15;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 DPXGLEHDXGPKG 16
DB 67 DPAGIRYGVGVGDK 81
I: |||
: |||

RESULT 7
D83818
heat-shock protein (activation of DnaK) dnaJ [imported] - Bacillus halodurans (strain
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: D83818
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-370 <STO>
A;Cross-references: GB:AP001511; GB:BA000004; NID:51017372; PIDN:BA05067.1; GSPDB:G
A;Experimental source: strain C-125
C;Genetics:

```

```

A43562
homeotic protein Hox D8 - mouse
N;Alternate names: homeotic protein Hox 4.3
C;Species: Mus musculus (house mouse)
C;Date: 11-Dec-1992 #sequence_revision 11-Dec-1992 #text_change 30-May-1997
C;Accession: A43562
R;Izpisua-Belmonte, J.C.; Dolle, P.; Renucci, A.; Zappavigna, V.; Falkenstein, H.; Du
Development 110, 733-745, 1990
A;Title: Primary structure and embryonic expression pattern of the mouse Hox-4.3 home
A;Reference number: A43562; MUID:91209232
A;Accession: A43562
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-289 <IZP>
A;Superfamily: homeotic protein Hox A7; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;196-252/Domain: homeobox homology <HOX>

Query Match 44.8%; Score 43; DB 2; Length 289;
Best Local Similarity 53.3%; Pred. No. 12;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 KDPXGLEHDXGPKG 15
DB 273 KEASGLEEDGAECCP 287
I: |||||
: |||||

RESULT 6
S42424
hypothetical protein Y - Mycobacterium smegmatis
C;Species: Mycobacterium smegmatis
C;Date: 07-May-1998 #sequence_revision 15-May-1998 #text_change 22-Oct-1999
C;Accession: S42424; S31804
R;Cirillo, J.D.; Weisbrod, T.R.; Pascopella, L.; Bloom, B.R.; Jacobs Jr., W.R.
Mol. Microbiol. 11, 629-639, 1994
A;Title: Isolation and characterization of the aspartokinase and aspartate semialdehy
A;Reference number: S42421; MUID:94254720
A;Accession: S42424
A;Molecule type: DNA
A;Residues: 1-333 <CIR>
A;Cross-references: EMBL:Z17372; NID:644506; PIDN:CAA78987.1; PID:9581353
A;Note: the authors translated the initiation codon GTG for residue 1 as Val
C;Genetics:
A;Start codon: GTG

Query Match 44.8%; Score 43; DB 2; Length 333;
Best Local Similarity 40.0%; Pred. No. 15;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 DPXGLEHDXGPKG 16
DB 67 DPAGIRYGVGVGDK 81
I: |||
: |||

RESULT 7
D83818
heat-shock protein (activation of DnaK) dnaJ [imported] - Bacillus halodurans (strain
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: D83818
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-370 <STO>
A;Cross-references: GB:AP001511; GB:BA000004; NID:51017372; PIDN:BA05067.1; GSPDB:G
A;Experimental source: strain C-125
C;Genetics:

```



Tue Jul 16 16:10:14 2002

Dev. Biol. 160, 413-423, 1993  
 A:Title: XFRH2, a Xenopus HNF-3 alpha homologue, exhibits both activin-inducible and  
 A:Reference number: I51580; MUID:94074768  
 A:Accession: I51580  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-427 <BOL>  
 A:Cross-references: GB:M93658; NID:g214901; PIDN:AAA17050.1; PID:g214902  
 C:Genetics:  
 A:Gene: XFKH2  
 C:Superfamily: transcription factor HNF-3; fork head DNA-binding domain homology  
 F:157-248/Domain: fork head DNA-binding domain homology <FHD>

Query Match 43.8%; Score 42; DB 2; Length 427;  
 Best Local Similarity 57.1%; Pred. No. 29;  
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 PXGLEHGDGNGXPKG 16  
 | | | | | | | | | |  
 Db 301 POALEHNGSNGEMK 314

## RESULT 15

WMBELM  
 membrane protein LMP-2A - human herpesvirus 4

N:Contains: membrane protein LMP-2B  
 C:Species: human herpesvirus 4, Epstein-Barr virus

A:Note: host Homo sapiens (man)  
 C:Date: 31-Dec-1989 #sequence\_revision 31-Dec-1989 #text\_change 16-Jul-1999

A:Accession: A30178; B30178; S00392

R:Sample: J.; Liebowitz, D.; Kieff, E.

J. Virol. 63, 933-937, 1989

A:Title: Two related Epstein-Barr virus membrane proteins are encoded by separate gen

A:Reference number: A30178; MUID:89095024

A:Accession: A30178

A:Molecule type: mRNA

A:Residues: 1-497 <SMA>  
 A:Cross-references: GB:M42412; NID:g522186; PIDN:AAA45887.1; PID:g522187

A:Accession: B30178

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 120-497 <SA2>

R:laux, G.; Perricaudet, M.; Farrell, P.J.

EMBO J. 7, 769-774, 1988

A:Title: A spliced Epstein-Barr virus gene expressed in immortalized lymphocytes is c

A:Reference number: S00392; MUID:88283646

A:Accession: S00392

A:Molecule type: mRNA

A:Residues: 1-497 <LAU>

A:Cross-references: EMBL:Y00835; NID:g59183; PIDN:CAA68762.1; PID:g59184

C:Genetics:

A:Gene: terminal

A:Introns: 140/2; 212/1; 245/1; 328/1; 355/1; 412/1; 484/1

C:Superfamily: Epstein-Barr virus membrane protein LMP-2A

C:Keywords: glycoprotein; transmembrane protein

F:120-497/Product: membrane protein LMP-2B #status predicted <LMB>

F:122-141/Domain: transmembrane #status predicted <TMA>

F:150-168/Domain: transmembrane #status predicted <TMB>

F:178-198/Domain: transmembrane #status predicted <TMC>

F:208-235/Domain: transmembrane #status predicted <TMD>

F:242-259/Domain: transmembrane #status predicted <TME>

F:267-288/Domain: transmembrane #status predicted <TMG>

F:300-316/Domain: transmembrane #status predicted <TMH>

F:321-339/Domain: transmembrane #status predicted <TMI>

F:355-373/Domain: transmembrane #status predicted <TMJ>

F:392-411/Domain: transmembrane #status predicted <TMK>

F:419-443/Domain: transmembrane #status predicted <TML>

F:450-470/Domain: transmembrane #status predicted <TML>

F:27,320,417/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match

43.8%; Score 42; DB 1; Length 497;

## RESULT 12

AH2441  
 hypothetical protein all5088 [imported] - Anabaena sp. (strain PCC 7120)

C:Species: Anabaena sp.

A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002

C:Accession: AH2441

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuchida

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.

NAKAZAKI, N.; SHIMPO, S.; SUGIMOTO, M.; TAKAZAWA, M.; YASUDA, M.; TABATA, S.

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AH2441

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-196 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA076787.1; PID:g17134226; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all5088

C:Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0226

Query Match 43.8%; Score 42; DB 2; Length 196;

Best Local Similarity 47.1%; Pred. No. 12;

Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 DPXGLEHGDGNGXPKGY 18

| | | | | | | | | |

Db 67 DDSGLQVDALNGLPGVY 83

## RESULT 13

H82320  
 HAM1 protein VC0456 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833

A:Accession: H82320

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-200 <HEI>

A:Cross-references: GB:AE004132; GB:AE003852; NID:g9654871; PIDN:AAF93629.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Map position: 1

C:Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0226

Query Match 43.8%; Score 42; DB 2; Length 200;

Best Local Similarity 52.9%; Pred. No. 12;

Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 DPXGLEHGDGNGXPKGY 18

| | | | | | | | | |

Db 68 DDSGLQVDALNGLPGVY 84

## RESULT 14

I51580  
 XFKH2 protein - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 16-Jul-1999

C:Accession: I51580

R:Polce, M.E.; Hemmati-Brivanlou, A.; Harland, R.M.

Tue Jul 16 16:10:14 2002

us-09-394-019a-248.rpr

Best Local Similarity 63.6%; Pred. No. 34;  
Matches 7; Conservative 2; Mismatches

2; Indels 0; Gaps 0;

QY 5 GLEHGDGNGXP 15

Db 87 GLQHDGNDGLP 97

Search completed: July 16, 2002, 11:20:52  
Job time: 330 sec

**THIS PAGE BLANK (copy)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2002, 11:20:22 ; Search time 10.36 Seconds  
(without alignments)  
67.273 Million cell updates/sec

Title: US-09-394-019a-248

Perfect score: 96

Sequence: 1 KDPXGLEHGDINGXPKGY 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43.5	45.3	655	1	ABG2_HUMAN
2	43	44.8	289	1	HXD8_MOUSE
3	43	44.8	333	1	YASD_MYCSM
4	43	44.8	352	1	NDR4_HUMAN
5	43	44.8	352	1	NDR4_RAT
6	42.5	44.3	370	1	DNAJ_BACHD
7	42	43.8	282	1	YC80_GUITH
8	42	43.8	375	1	NDR3_MOUSE
9	42	43.8	427	1	FKH2_XENLA
10	42	43.8	497	1	LMP2_EBV
11	42	43.8	724	1	TAP1_MOUSE
12	42	43.8	725	1	TAP1_RAT
13	41.5	43.2	327	1	FXA2_SALTI
14	41.5	43.2	327	1	FXA2_SALTY
15	41	42.7	448	1	PU17_DROME
16	41	42.7	798	1	UNR_HUMAN
17	41	42.7	798	1	UNR_RAT
18	41	42.7	1182	1	ABL2_HUMAN
19	40.5	42.2	1451	1	DPO3_MYCGE
20	40	41.7	375	1	PGK_TETPY
21	40	41.7	481	1	BIND_STRPU
22	39.5	41.1	397	1	YEDS_ECOLI
23	39	40.6	289	1	DLX5_MOUSE
24	39	40.6	471	1	DNAB_SALTY
25	39	40.6	583	1	MLIX_MOUSE
26	39	40.6	787	1	SYFB_NEIMA
27	39	40.6	787	1	SYFB_NEIMB
28	38	39.6	309	1	YT92_RHILO
29	38	39.6	379	1	YJHT_HAEIN
30	38	39.6	459	1	DNAB_BUCAI
31	38	39.6	471	1	DNAB_ECOLI
32	38	39.6	525	1	HRG_HUMAN
33	38	39.6	574	1	KPYR_RAT

34	38	39.6	1763	1	POLN_FCVF9
35	37.5	39.1	650	1	ABG3_MOUSE
36	37.5	39.1	922	1	PMP1_CHLPN
37	37	38.5	136	1	GLB3_CHITP
38	37	38.5	151	1	GLB3_CHITH
39	37	38.5	151	1	GLB4_CHITH
40	37	38.5	222	1	VP32_YEAST
41	37	38.5	248	1	HMD1_CHICK
42	37	38.5	302	1	LGUL_SCHPO
43	37	38.5	334	1	CHMU_ARATH
44	37	38.5	352	1	NIVO_CLOPA
45	37	38.5	356	1	MPCP_RAT

P27409 feline calli  
Q99p81 mus musculus  
Q929d5 chlamydia p  
P22431 chironomus  
P02229 chironomus  
P02230 chironomus  
P51996 saccharomyc  
P46692 gallus galli  
Q09751 schizosacch  
P42738 arabidopsis  
Q00852 clostridium  
P16036 rattus norv

## ALIGNMENTS

RESULT 1  
ABG2\_HUMAN  
ID ABG2\_HUMAN STANDARD: PRT; 655 AA.  
AC Q9UNQ0; Q95374; Q9NUS0; Q9BY73;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DE ATP-binding cassette, sub-family G, member 2 (Placenta-specific ATP-binding cassette transporter) (Breast cancer resistance protein).  
GN ABCG2 OR ABCP OR BCRP OR BCRP1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=99065313; PubMed=9850061;  
RA Allikmets R., Schriml L.M., Hutchinson A., Romano-Spica V., Dean M.;  
RT "A human placenta-specific ATP-binding cassette gene (ABCP) on chromosome 4q22 that is involved in multidrug resistance.";  
RL Cancer Res. 58:5337-5339(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Breast cancer;  
RX MEDLINE=99080071; PubMed=9861027;  
RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,  
RS Ross D.D.;  
RT "A multidrug resistance transporter from human MCF-7 breast cancer cells.";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:15665-15670(1998).  
RN [3]  
RP ERRATUM.  
RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,  
RS Ross D.D.;  
RT "Breast cancer resistance protein constitutes a 140-kDa complex as a homodimer.";  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 198-655 FROM N.A.  
RC TISSUE=Placenta;  
RA Itoigai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
RA Wagatsuma M., Hosokawa T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Yamamoto J., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Ninomiya K., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]

```

Query Match      45.3%; Score 43.5; DB 1; Length 655;
Best Local Similarity 58.8%; Pred. NO. 14;
Matches, 10; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

          1 KDPXGLEHDG-INGXPK 16
          ||| ||| | ||| |
          97 KDPSGLSGDVLINGAPR 113

RESULT 2
HYXD8_MOUSE|
IID HXD8_MOUSE STANDARD; PRT; 289 AA.
AAC P23463;
01-NOV-1991 (Rel. 20, Created)
DDT 01-JUL-1993 (Rel. 26, Last sequence update)

```



FT DOMAIN 183 188 ANTP-TYPE HEXAPEPTIDE.  
 FT DNA\_BIND 195 254 HOMEBOX.  
 FT CONFLICT 207 208 TL -> RV (IN REF. 1).  
 FT CONFLICT 231 231 T -> S (IN REF. 1).  
 FT CONFLICT 265 266 EA -> DG (IN REF. 1).  
 FT CONFLICT 275 275 A -> V (IN REF. 2).  
 SQ SEQUENCE 289 AA; 31410 MW; 5783099FB9B2BDFE CRC64;

Query Match 44.8%; Score 43; DB 1; Length 289;  
 Best Local Similarity 53.3%; Pred. No. 6.8;  
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 KDPXGLEHDGNGXP 15  
 I: ||| || I I  
 Db 273 KEASGLEEDGAECCP 287

RESULT 3  
 YASD\_MYCSM  
 ID YASD\_MYCSM STANDARD; PRT; 333 AA.  
 AC P41402;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 35.9 kDa protein in ASD 3'region (ORFV).  
 OS Mycobacterium smegmatis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 NCBI\_TaxID=1772;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=ATCC 607 / MC(2)6 / NRRL B-692;  
 RX MEDLINE=94254720; PubMed=7910936;  
 RA Cirillo J.D., Weissbrod T.R., Pascopella L., Bloom B.R.,  
 RA Jacobs W.R., Jr.;  
 RT "Isolation and characterization of the aspartokinase and aspartate  
 RT semialdehyde dehydrogenase operon from mycobacteria";  
 RL Mol. Microbiol. 11:629-639 (1994).  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: Z17372; CAA78987.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 333 AA; 35881 MW; C1C50D9A6DC37368 CRC64;

Query Match 44.8%; Score 43; DB 1; Length 333;  
 Best Local Similarity 40.0%; Pred. No. 8;  
 Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
 QY 2 DPXGLEHDGNGXPK 16  
 I: || I: ::||: I I  
 Db 67 DPAGIRNGVSGVDK 81

RESULT 4  
 NDRA\_HUMAN  
 ID NDRA4\_HUMAN STANDARD; PRT; 352 AA.  
 AC Q9ULP0; Q9GZX0; Q9GZN3; Q9GZM1;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DE NDRA4 protein (Brain development-related molecule 1) (Vascular smooth  
 DE muscle cell associated protein-8) (SMAP-8).  
 GN NDRA4 OR BDW1 OR KIAA1180.  
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.  
 RX MEDLINE=21251006; PubMed=11352569;  
 RA Zhou R.-H., Kokame K., Tsukamoto Y., Yutani C., Kato H., Miyata T.;  
 RT "Characterization of the human NDRG gene family: a newly identified  
 RT member, NDRG4, is specifically expressed in brain and heart";  
 RL Genomics 73:86-97 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 3).  
 RC TISSUE=Heart;  
 RA Nishimoto S., Tawara J., Aoki T., Toyoda H., Komuraaki T.;  
 RT "Molecular cloning and characterization of the human vascular smooth  
 RT muscle cell associated protein-8 (SMAP-8)";  
 RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Amygdala;  
 RX MEDLINE=21154917; PubMed=11230166;  
 RA Wismann S., Weill B., Weilenreuther R., Gassenhuber J., Glassl S.,  
 RA Ansoerge W., Boecker M., Bloeker H., Bauersachs S., Blum H.,  
 RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,  
 RA Meves H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,  
 RA Wambutt R., Korn B., Klein M., Poustka A.;  
 RT "Towards a catalog of human genes and proteins: sequencing and  
 RT analysis of 500 novel complete protein coding human cDNAs";  
 RL Genome Res. 11:422-435 (2001).  
 RN [4]  
 RP SEQUENCE OF 80-352 FROM N.A. (ISOFORM 1).  
 RC TISSUE=Brain;  
 RX MEDLINE=20039618; PubMed=10574461;  
 RA Hiroseawa M., Nagase T., Ishikawa K.-I., Kikuno R., Nomura N.,  
 RA Ohara O.;  
 RT "Characterization of cDNA clones selected by the GeneMark analysis  
 RT from size-fractionated cDNA libraries from human brain";  
 RL DNA Res. 6:329-336 (1999).  
 CC -!- FUNCTION: MAY PLAY A ROLE IN THE EARLY POSTNATAL DEVELOPMENT AND  
 CC FUNCTION OF NEURONAL CELLS (BY SIMILARITY).  
 CC -!- ALTERNATIVE PRODUCTS: 3 isoforms; 1/NDRG4-Bvar (shown here),  
 CC 2/NDRG4-B and 3/NDRG4-H; are produced by alternative splicing.  
 CC -!- TISSUE SPECIFICITY: Expressed only in brain and heart. Isoforms 1  
 CC and 2 are only expressed in brain. Isoform 3 is expressed in both  
 CC heart and brain.  
 CC -!- SIMILARITY: BELONGS TO THE NDRG FAMILY.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: AB044947; BAB20071.1; -  
 DR EMBL: AB044947; BAB20072.1; -  
 DR EMBL: AB044947; BAB20073.1; -  
 DR EMBL: AB044944; BAB20068.1; -  
 DR EMBL: AB044945; BAB20069.1; -  
 DR EMBL: AB044946; BAB20070.1; -  
 DR EMBL: AB021172; BAB20288.1; -  
 DR EMBL: AL136584; CAB66519.1; -  
 DR EMBL: AB033006; BAA86494.1; -  
 DR InterPro: IPR004142; Ndr.  
 DR Pfam: PF03096; Ndr; 1.  
 KW Developmental protein; Alternative splicing.  
 FT DOMAIN 253 256 POLY-THR.  
 FT VARSPLIC 1 7 MPECWDG -> MAGLOELRPPEKPLLRQDATELESSDAF  
 FT VARSPLIC 289 302 LLAADTDWK (IN ISOFORM 3).  
 FT IATLKRRRSGGAV -> M (IN ISOFORM 2 AND  
 FT ISOFORM 3).  
 SQ SEQUENCE 352 AA; 38459 MW; 4CFAC7C820013B6 CRC64;

Query Match	44.8%	Score 43;	DB 1;	Length 352;
Best Local Similarity	60.0%	Pred. NO.	8 5;	
Matches	9;	Conservative	0;	Mismatches 6; Indels 0; Gaps 0;
QY	3	PXGLEHGDINGXPKG	17	
Db	14	PYGLLHVIVRGSPKG	28	
RESULT	5			
ID	NDR4_RAT	STANDARD;	PRT;	352 AA.
AC	Q9Z2L9;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	NDRG4 protein (Brain development-related molecule 1).			
GN	NDRG4 Or BDM1.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RX	MEDLINE=99255674; PubMed=10320792;			
RA	Yamauchi Y., Hongo S., Nishinaka N., Ohashi T., Takahashi R.,			
RA	Takeda M.; Characterization of a novel developmentally			

```
RT "Molecular cloning and characterization of the rat brain-specific gene Bdm1, showing predominant expression in postnatal rat brain.";
RL Brain.Res. Mol. Brain Res. 68:149-158(1999).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE EARLY POSTNATAL DEVELOPMENT AND
CC FUNCTION OF NEURONAL CELLS.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE BRAIN AND HEART, WEAKLY IN
CC THE KIDNEY; MOST PROMINENTLY IN POSTNATAL BRAIN WHERE IT IS
CC EXPRESSED WIDELY IN THE OLFACTORY BULB, CEREBRAL CORTEX,
CC HIPPOCAMPUS, CEREBELLUM, THALAMUS AND MEDULLA OBLONGATA.
CC -!- SIMILARITY: BELONGS TO THE NDRG FAMILY.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AF045564; AA02415.1; -.
DR InterPro; IPR004142; Ndr.
DR Pfam; PF03096; Ndr; 1.
KW Developmental protein.
FT DOMAIN 253 256 POLY-THR.
FT SEQUENCE 352 AA; 38487 MW; EF35D773B9D255D9 CRC64;
-----
Query Match 44.8%; Score 43; DB 1; Length 352;
Best Local similarity 60.0%; Pred. No. 8.5; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 6;
QY 3 |PXGLEHGDINGXPKG 17
DB 14 |PYGLLHVIVIRGSPKG 28
| | | | | | | |
RESULT 6
DNAJ_BACHD STANDARD; PRT; 370 AA.
ID DNAJ_BACHD
AC Q9K071;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
```

```
RESULT 7
YC80_GUITH          STANDARD;          PRT;      282 AA.
AC 078449;
AT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Hypothetical 33.2 kDa protein ycf80.
GN YCF80.
OS Guillardia theta (Cryptomonas phi).
OC Chloroplast.
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX NCBI_TaxID=55529;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99128221; PubMed=9929392;
RA Douglas S.E., Penny S.L.;
RT "The plastid genome of the cryptophyte alga, Guillardia theta:
RT complete sequence and conserved syntenic groups confirm its common
RT ancestry with red algae.";
CC J. Mol. Evol. 48:236-244(1999).
CC -!- SIMILARITY: BELONGS TO THE YCF80 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF041468; AAC35638.1; -
KW Hypothetical protein; Chloroplast.
SQ SEQUENCE 282 AA; 33221 MW; 0D84447DCADA943A CRC64;

Query Match          43.8%; Score 42; DB 1; Length 282;
Best Local Similarity 58.3%; Pred. No. 9.7;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 KDPXGLEHDGIN 12
   ||| ||| |||
DB 175 KDPGCVGVN 186

RESULT 8
NDR3_MOUSE
ID NDR3_MOUSE          STANDARD;          PRT;      375 AA.
AC Q90YF9;
AT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE NDR3 protein (Ndr3 protein).
GN NDR3 OR NDR3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20050077; PubMed=10581191;
RA Okuda T., Kondoh H.;
RT "Identification of new genes ndr2 and ndr3 which are related to
RT Ndr1/RTP/Drg1 but show distinct tissue specificity and response to
RT N-myc.";
RL Biochem. Biophys. Res. Commun. 266:208-215(1999).
CC -!- TISSUE SPECIFICITY: THYMUS, NERVOUS SYSTEM, GUT AND KIDNEY.
CC -!- DEVELOPMENTAL STAGE: ITS EXPRESSION IS ALREADY SIGNIFICANT AT 9.5
CC DPC, COVERING THE ENTIRE EMBRYO EXCEPT THE HEART, AND IT SHOWS
CC ONLY A SLIGHT INCREASE IN LATER DEVELOPMENTAL STAGES.
CC -!- SIMILARITY: BELONGS TO THE NDRG FAMILY.
CC -----
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB033922; BAA85883.1; -
DR MGD; MGI:1352499; Ndr3.
DR InterPro; IPR000379; Est_lip_thioest_actsite.
DR InterPro; IPR004142; Ndr.
DR Pfam; PF03096; Ndr; 1.
SQ SEQUENCE 375 AA; 41555 MW; B33D3CC3E816AE1A CRC64;

Query Match          43.8%; Score 42; DB 1; Length 375;
Best Local Similarity 53.3%; Pred. No. 13;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 3 PXGLEHDGNGXPKG 17
   | | | | | | | |
DB 38 PHGMVHTVIRGLPKG 52

RESULT 9
FKH2_XENLA
ID FKH2_XENLA          STANDARD;          PRT;      427 AA.
AC P32315;
AT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE XFKH2 protein.
GN XFKH2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94074768; PubMed=8253274;
RA Bolce M.E., Hemmati-Briylanlou A., Harland R.M.;
RT "XFKH2, a Xenopus HNF-3 alpha homologue, exhibits both
RT activin-inducible and autonomous phases of expression in early
RT embryos.";
RL Dev. Biol. 160:413-423(1993).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- TISSUE SPECIFICITY: PRESENT IN THE VEGETAL POLE AND MARGINAL ZONE
CC BUT NOT THE ANIMAL POLE OF GASTRULAE AND IN EQUAL LEVELS IN THE
CC DORSAL AND VENTRAL HALVES OF BOTH GASTRULAE AND NEURULAE.
CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M93658; AAA17050.1; -
DR HSSP; Q63245; 2HFP.
DR InterPro; IPR001766; Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS00039; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear protein.
FT DNA_BIND 156 247
SQ SEQUENCE 427 AA; 46572 MW; 2D29A42AF60730C CRC64;
```

Tue Jul 16 16:10:14 2002

us-09-394-019a-248.rsp

```

Query Match      43.8%; Score 42; DB 1; Length 427;
Best Local Similarity 57.1%; Pred. No. 15;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 PXGLEHGGINGXPK 16
   | |||:| | | |
Db 301 PQALEHNGSNGEMK 314

RESULT 10
LMP2_EBV
ID LMP2_EBV STANDARD; PRT; 497 AA.
AC P13255;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Gene terminal protein (Membrane protein LMP-2A/LMP-2B).
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88283646; PubMed=2840285;
RA Laux G., Perricaudet M., Farrell P.J.;
RT "A spliced Epstein-Barr virus gene expressed in immortalized
   lymphocytes is created by circularization of the linear viral
   genome.";
RL EMO J. 7:769-774(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89095024; PubMed=2536113;
RA Sample J., Liebowitz D., Kieff E.;
RT "Two related Epstein-Barr virus membrane proteins are encoded by
   separate genes.";
RL J. Virol. 63:933-937(1989).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tufnell P.S., Barrell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
CC -!- FUNCTION: MAY BE IMPORTANT FOR IMMORTALIZATION OF CELLS BY EBV.
CC -!- SUBCELLULAR LOCATION: MEMBRANE (PROBABLE).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24212; AAA45887.1; -
CC EMBL; Y00835; CAA68762.1; -
CC EMBL; V01555; -; NOT_ANNOTATED_CDS.
CC PIR; A30178; WNBELM.
KW Transmembrane.
FT CHAIN 1 497 MEMBRANE PROTEIN LMP-2A.
FT CHAIN 120 497 MEMBRANE PROTEIN LMP-2B.
FT TRANSMEM 122 141 POTENTIAL.
FT TRANSMEM 150 168 POTENTIAL.
FT TRANSMEM 178 198 POTENTIAL.
FT TRANSMEM 208 235 POTENTIAL.
FT TRANSMEM 242 259 POTENTIAL.
FT TRANSMEM 267 288 POTENTIAL.
FT TRANSMEM 300 316 POTENTIAL.
FT TRANSMEM 321 339 POTENTIAL.
FT TRANSMEM 355 373 POTENTIAL.

Query Match      43.8%; Score 42; DB 1; Length 497;
Best Local Similarity 63.6%; Pred. No. 18;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 GLEHGGINGXP 15
   |||:| | |
Db 87 GLQHDGNDGLP 97

RESULT 11
TAP1_MOUSE
ID TAP1_MOUSE STANDARD; PRT; 724 AA.
AC P21958;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigen peptide transporter 1 (APT1) (Histocompatibility antigen
   modifier 1).
DE ABCB2 OR TAP1 OR HAM-1.
GN ABCB2 OR TAP1 OR HAM-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=WEHI-3;
RX MEDLINE=97307601; PubMed=9164943;
RA Marusina K., Iyer M., Monaco J.J.;
RT "Allelic variation in the mouse Tap-1 and Tap-2 transporter genes.";
RL J. Immunol. 158:5251-5256(1997).
RN [2]
RP SEQUENCE OF 148-724 FROM N.A.
RX MEDLINE=91102550; PubMed=2270487;
RA Monaco J.J., Cho S., Attaya M.;
RT "Transport protein genes in the murine MHC: possible implications for
   antigen processing.";
RL Science 250:1723-1726(1990).
RN [3]
RP SEQUENCE OF 148-724 FROM N.A.
RX STRAIN=BALB/C;
RA Sun X.Y., Zhou J., Frazer I.;
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE TRANSPORT OF ANTIGENS FROM THE CYTOPLASM
CC TO A MEMBRANE-BOUND COMPARTMENT FOR ASSOCIATION WITH MHC CLASS I
CC MOLECULES.
CC -!- SUBUNIT: HETERODIMER OF TAP1 AND TAP2.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U60018; ABA41962.1; -
CC EMBL; M55637; AAA39570.1; -
CC EMBL; X59615; CAA42178.1; -
CC PIR; A37779; A37779.
CC MGD; MGI:98483; Abcb2.
CC InterPro; IPR003593; AAA.
CC InterPro; IPR001140; ABC_transporter_tmnm.
CC InterPro; IPR003439; ABC_transportr.
CC InterPro; IPR001687; ATP_GTP_A.
CC Pfam; PF00664; ABC_membrane; 1.

```

DR Pfam; PF00005; ABC\_tran; 1.  
DR SMART; SM00382; AAA; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
KW Peptide transport; Transmembrane; ATP-binding.  
FT TRANSMEM 69 POTENTIAL.  
FT TRANSMEM 116 136 POTENTIAL.  
FT TRANSMEM 163 183 POTENTIAL.  
FT TRANSMEM 204 224 POTENTIAL.  
FT TRANSMEM 289 309 POTENTIAL.  
FT TRANSMEM 391 411 POTENTIAL.  
FT TRANSMEM 420 440 POTENTIAL.  
FT NP\_BIND 514 521 ATP (POTENTIAL).  
FT CONFLICT 355 355 S -> Q (IN REF. 3).  
FT CONFLICT 481 481 F -> L (IN REF. 3).  
SQ SEQUENCE 724 AA; 79005 MW; 56A50E6ED4C89CAF CRC64;

Query Match 43.8%; Score 42; DB 1; Length 724;  
Best Local Similarity 63.6%; Pred. No. 27;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 8 HDGINGXPKGY 18  
|||:|:|:|:  
DB 598 HDFISGFPGY 608

RESULT 12  
TAP1\_RAT  
ID TAP1\_RAT STANDARD; PRT; 725 AA.  
AC P36370.  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Antigen peptide transporter 1 (Apt1).  
GN ABCB2 OR TAP1 OR MTPI.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91080926; PubMed=1979660;  
RA Deverson E.V., Gow I.R., Coadwell W.J., Monaco J.J., Butcher G.W.,  
RA Howard J.C.;  
RT "MHC class II region encoding proteins related to the multidrug  
resistance family of transmembrane transporters.";  
RL Nature 348:738-741(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BDIX; TISSUE=Lymphocytes;  
RA Deverson E.V.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: INVOLVED IN THE TRANSPORT OF ANTIGENS FROM THE CYTOPLASM  
TO A MEMBRANE-BOUND COMPARTMENT FOR ASSOCIATION WITH MHC CLASS I  
MOLECULES.  
CC -1- SUBUNIT: HETERODIMER OF TAP1 AND TAP2.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X57523; CAA40742.1; ALT\_INIT.  
DR EMBL; Y10231; CAA71280.1;  
DR InterPro; IPR003593; AAA.  
DR InterPro; IPR001140; ABC\_transporter\_tmem.  
DR InterPro; IPR003439; ABC\_transporter.  
DR InterPro; IPR001687; ATP\_GTP\_A.

DR Pfam; PF00664; ABC\_membrane; 1.  
DR SMART; SM00382; AAA; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER; 1.  
KW Peptide transport; Transmembrane; ATP-binding.  
FT TRANSMEM 68 POTENTIAL.  
FT TRANSMEM 117 137 POTENTIAL.  
FT TRANSMEM 164 184 POTENTIAL.  
FT TRANSMEM 205 225 POTENTIAL.  
FT TRANSMEM 298 318 POTENTIAL.  
FT TRANSMEM 392 412 POTENTIAL.  
FT TRANSMEM 421 441 POTENTIAL.  
FT NP\_BIND 515 522 ATP (POTENTIAL).  
FT CARBOHYD 227 227 N-LINKED (GLCNAC..) (POTENTIAL).  
SQ SEQUENCE 725 AA; 79150 MW; 3FA7215D0AC22EE0 CRC64;

Query Match 43.8%; Score 42; DB 1; Length 725;  
Best Local Similarity 63.6%; Pred. No. 27;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 8 HDGINGXPKGY 18  
|||:|:|:|:  
DB 599 HDFISGFPGY 609

RESULT 13  
PXA2\_SALTI  
ID PXA2\_SALTI STANDARD; PRT; 327 AA.  
AC P58716;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE 4-hydroxythreonine-4-phosphate dehydrogenase 2 (EC 1.1.1.262) (4-  
(phosphohydroxy)-L-threonine dehydrogenase 2).  
GN PXA2 OR STV0185.  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatia M.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
RA Cronin A., Davies P., Davies R.M., Dowd L., White N., Farrar J.,  
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,  
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrrell B.G.;  
RL "Complete genome sequence of a multiple drug resistant Salmonella  
enterica serovar Typhi CT18.";  
RL Nature 413:848-852(2001).  
CC -1- FUNCTION: Catalyzes the NAD-dependent oxidation of 4-  
(phosphohydroxy)-L-threonine (HTP) into 2-amino-3-oxo-4-  
(phosphohydroxy)butyric acid which spontaneously decarboxylate to  
form L-amino-3-(phosphohydroxy)propan-2-one (3-amino-2-oxopropyl  
phosphate) (By similarity).  
CC -1- CATALYTIC ACTIVITY: 4-(phosphonoxy)-threonine + NAD(+) = 2-  
amino-3-oxo-4-phosphonoxybutyrate + NADH.  
CC -1- PATHWAY: De novo synthesis of pyridoxine (vitamin B6) and  
pyridoxal phosphate.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -1- SIMILARITY: BELONGS TO THE PXA FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC

```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL627265; CAD01321.1;
CC Pyridoxine biosynthesis; Oxidoreductase; NAD; Complete proteome.
CC KW Pyridoxine biosynthesis; Oxidoreductase; NAD; Complete proteome.
CC SQ SEQUENCE 327 AA; 35041 MW; F7F563DEB5326FEB CRC64;

Query Match 43.2%; Score 41.5; DB 1; Length 327;
Best Local Similarity 39.1%; Pred. NO. 14;
Matches 9; Conservative 1; Mismatches 4; Indels 9; Gaps 1;

Qy 2 DPXG-----LEHGDINGXP 15
||| | | | | |
Db 12 DPAGIGPEIIIVKALSEDGLNGAP 34

RESULT 14
FXA2_SALTY STANDARD; PRT; 327 AA.
ID P58718;
AC 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE 4-hydroxythreonine-4-phosphate dehydrogenase 2 (EC 1.1.1.262) (4-
DE (phosphonyldioxy)-L-threonine dehydrogenase 2).
DE PDXA2 OR STM0163.
GN Salmonella typhimurium.
OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OC NCBI_TaxID=602;
OX [1]
RN SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=2153498; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Powollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Watson R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
CC -!- FUNCTION: Catalyzes the NAD-dependent oxidation of 4-
CC (phosphonyldioxy)-L-threonine (Htp) into 2-amino-3-oxo-4-
CC (phosphonyldioxy)butyric acid which spontaneously decarboxylate to
CC form 1-amino-3-(phosphonyldioxy)propan-2-one (3-amino-2-oxopropyl
CC phosphate) (by similarity).
CC -!- CATALYTIC ACTIVITY: 4-(phosphonyldioxy)-threonine + NAD(+) = 2-
CC amino-3-oxo-4-phosphonyldioxybutyrate + NADH.
CC -!- PATHWAY: De novo synthesis of pyridoxine (vitamin B6) and
CC pyridoxal phosphate.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: BELONGS TO THE PDXA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB008701; AAL19127.1;
CC StyGene: SG272727; pdxA2.
CC KW Pyridoxine biosynthesis; Oxidoreductase; NAD; Complete proteome.
CC SQ SEQUENCE 327 AA; 35064 MW; 03BB6725F1896440 CRC64;

Query Match 43.2%; Score 41.5; DB 1; Length 327;
Best Local Similarity 39.1%; Pred. NO. 14;
Matches 9; Conservative 1; Mismatches 4; Indels 9; Gaps 1;

Qy 2 DPXG-----LEHGDINGXP 15
||| | | | | |

```

```

Db 12 DPAGIGPEIIIVKALSEDGLNGAP 34
||| | | | | |
RESULT 15
RUI7_DROME STANDARD; PRT; 448 AA.
ID RUI733; OQVMS56;
AC 01-AUG-1990 (Rel. 15, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE U1 small nuclear ribonucleoprotein 70 kDa (U1 snRNP 70 kDa) (snRNP70).
GN SNRNP70K OR SNRNP27D OR CG8749.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90258833; PubMed=1692955;
RA Mancebo R., Lo P.C.H., Mount S.M.;
RA "Structure and expression of the drosophila melanogaster gene for the
RT U1 small nuclear ribonucleoprotein particle 70K protein."
RL Mol. Cell. Biol. 10:2492-2502(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkelley;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bencan P.V., Bernan B.P., Bhandari D., Brottier P.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: MEDIATES THE SPLICING OF PRE-MRNA BY BINDING TO THE STEM
CC LOOP I REGION OF U1-SNRNA.
CC -!- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL627265; CAD01321.1;
CC Pyridoxine biosynthesis; Oxidoreductase; NAD; Complete proteome.
CC KW Pyridoxine biosynthesis; Oxidoreductase; NAD; Complete proteome.
CC SQ SEQUENCE 327 AA; 35041 MW; F7F563DEB5326FEB CRC64;

```

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M31162; AAA28859.1; -  
DR EMBL; AE003615; AAF52471.1; -  
DR PIR; A36311; A36311.  
DR HSP; P09651; IHA1.  
DR FlyBase; FBgn0016978; snRNP70K.  
DR InterPro; IPR000504; RRM.  
DR Pfam; PF00076; rrm; 1.  
DR SMART; SM00360; RRM; 1.  
DR PROSITE; PS50102; RRM; 1.  
DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
KW Nuclear protein; Ribonucleoprotein; RNA-binding; mRNA processing.  
FT DOMAIN 102 180 RNA-BINDING (RRM).  
FT DOMAIN 254 350 ARG/GLU-RICH (MIXED CHARGE).  
FT CONFLICT 278 278 N -> S (IN REF. 1).  
SQ SEQUENCE 448 AA; 52900 MW; 0DDFB5A39CA72AEB CRC64;

Query Match 42.7%; Score 41; DB 1; Length 448;  
Best Local Similarity 61.5%; Pred. No. 24;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 6 LEHDGNGXPKGY 18  
| | | | |  
Db 133 LIHQESGKPKGY 145

Search completed: July 16, 2002, 11:22:17  
Job time: 115 sec

**THIS PAGE BLANK (33PT0)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2002, 11:16:57 ; Search time 25.08 seconds  
(without alignments)  
124.159 Million cell updates/sec

Title: US-09-394-019a-248  
Perfect score: 96  
Sequence: 1 KDPXGLEHNGXPKGY 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 08  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phage.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*
  - 15: sp\_rviro.\*
  - 16: sp\_bacteriap.\*
  - 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	50.0	590	2 Q93CA0	Q93ca0 bifidobacte
2	47	49.0	451	5 Q9VPG0	Q9vpg0 drosophila
3	45	46.9	135	16 Q9A3D9	Q9a3d9 caulobacter
4	44.5	46.4	117	12 Q69133	Q69133 human herpe
5	44.5	46.4	118	12 Q66567	Q66567 human herpe
6	44.5	46.4	118	12 Q66568	Q66568 human herpe
7	44.5	46.4	118	12 Q66570	Q66570 human herpe
8	44.5	46.4	118	12 Q66573	Q66573 human herpe
9	44.5	46.4	118	12 Q66550	Q66550 human herpe
10	44.5	46.4	118	12 Q66551	Q66551 human herpe
11	44.5	46.4	118	12 Q66552	Q66552 human herpe
12	44.5	46.4	118	12 Q66553	Q66553 human herpe
13	44.5	46.4	118	12 Q66555	Q66555 human herpe
14	44.5	46.4	118	12 Q66556	Q66556 human herpe
15	44.5	46.4	118	12 Q66557	Q66557 human herpe
16	44.5	46.4	118	12 Q66559	Q66559 human herpe

17	44.5	46.4	118	12 Q66560	Q66560 human herpe
18	44.5	46.4	118	12 Q66561	Q66561 human herpe
19	44.5	46.4	118	12 Q66562	Q66562 human herpe
20	44.5	46.4	118	12 Q66563	Q66563 human herpe
21	44.5	46.4	118	12 Q66565	Q66565 human herpe
22	44.5	46.4	183	12 Q69134	Q69134 human herpe
23	44.5	46.4	542	16 Q9X287	Q9x287 thermotoga
24	44.5	46.4	826	5 Q9B130	Q9b130 ciona intes
25	43.5	45.3	655	4 Q96TA8	Q96ta8 homo sapien
26	43.5	45.3	655	4 Q96LD6	Q96ld6 homo sapien
27	43.5	45.3	1146	2 Q68317	Q68317 vibrio chol
28	43.5	45.3	1146	16 Q9KRT8	Q9krt8 vibrio chol
29	43	44.8	339	4 Q9GZM3	Q9gzm3 homo sapien
30	43	44.8	339	4 Q96PL9	Q96pl9 homo sapien
31	43	44.8	339	11 Q923D7	Q923d7 mus musculu
32	43	44.8	352	4 Q9GZM1	Q9gzml homo sapien
33	43	44.8	371	4 Q9GZM1	Q9gzml homo sapien
34	42.5	44.3	183	5 Q9VJ96	Q9vj96 drosophila
35	42	43.8	73	6 Q28751	Q28751 ovls sp. ke
36	42	43.8	118	12 Q66566	Q66566 human herpe
37	42	43.8	118	12 Q66569	Q66569 human herpe
38	42	43.8	118	12 Q66571	Q66571 human herpe
39	42	43.8	118	12 Q66572	Q66572 human herpe
40	42	43.8	118	12 Q66574	Q66574 human herpe
41	42	43.8	118	12 Q66554	Q66554 human herpe
42	42	43.8	118	12 Q66558	Q66558 human herpe
43	42	43.8	129	11 Q63911	Q63911 mus sp. tap
44	42	43.8	200	16 Q9KU09	Q9kuq9 vibrio chol
45	42	43.8	211	4 Q9BY23	Q9by23 homo sapien

ALIGNMENTS

RESULT 1

Q93CA0 PRELIMINARY; PRT; 590 AA.

ID Q93CA0; PRELIMINARY; PRT; 590 AA.

AC Q93CA0; PRELIMINARY; PRT; 590 AA.

DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE ALPHEA-GLUCOSIDASE.

GN AGLB.

OS Bifidobacterium adolescentis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium.

OX NCBI\_TaxID=1680;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=DSM 20083.

RA Van den Broek L.A.M., Struijs K., Voragen A.G.J., Verdoes J.C.,

RA Belman G.;

RT "Cloning and characterization of two alpha-glucosidases from

RT Bifidobacterium adolescentis.";

RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF411186; AAL05573.1;

SQ SEQUENCE 590 AA; 66575 MW; 545933DDC93B3CEB CRC64;

Query Match 50.0%; Score 48; DB 2; Length 590;

Best Local Similarity 52.9%; Pred. No. 8.9;

Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 KDPXGLEHNGXPKG 17

Db 521 EDQPGKANDGANGPFGG 537

RESULT 2

Q9VPG0 PRELIMINARY; PRT; 451 AA.

ID Q9VPG0; PRELIMINARY; PRT; 451 AA.

AC Q9VPG0; PRELIMINARY; PRT; 451 AA.

DT 01-MAY-2000 (TREMBLrel. 13, Created)

Tue Jul 16 16:10:14 2002

```

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CG5282 PROTEIN.
GN Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya P., Brothier P.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Cantler A., Chandra I.,
RA Burdick K.C., Busam D.A., Butler J., Dahlke C., Davenport L.B., Davies P.,
RA Cherry J.M., Cawley S., Dahlke C., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA de Pablos B., Delcher A., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fouts D.E., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA "the genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
RL EMBL; AE003591; AAF51593.1; -.
DR MEROPS; M19.UNW; -.
DR FlyBase; FBgn0036986; CG5282.
DR InterPro; IPR000180; Renal_dipeptidase.
DR Pfam; PF01244; Renal_dipeptidase; 1.
SQ SEQUENCE 451 AA; 49262 MW; 95103A0674F25C0F CRC64;

Query Match 49.08; Score 47; DB 5; Length 451;
Best Local Similarity 50.0%; Pred. No. 9.8;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 5 GLEHGDGNGXPKGY 18
DB 375 GVDHGLGLGAPKSY 388

RESULT 3
ID Q9A3D9 PRELIMINARY; PRT; 135 AA.
AC Q9A3D9;
DR 01-JUN-2001 (TReMBLrel. 17, Created)

DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN CC3265.
GN Caulobacter crescentus.
OS Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Niernan W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R., Ely B.,
RA Deboy C.I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Kolonay J.F., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Utterback T., Tran K., Wolf A., Yamathavan J., Berry K.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005990; AAK25227.1; -.
DR TIGR; CC3265; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 135 AA; 14607 MW; 01C21B2C6B498805 CRC64;

Query Match 46.9%; Score 45; DB 16; Length 135;
Best Local Similarity 50.0%; Pred. No. 5.7;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 5 GLEHGDGNGXPKGY 18
DB 89 GMDFDGDKGLPKGW 102

RESULT 4
ID Q69133 PRELIMINARY; PRT; 117 AA.
AC Q69133;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE LATENT MEMBRANE PROTEIN 2A (FRAGMENT).
GN LMP 2A.
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10376;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88283646; PubMed=2840285;
RA Laux G., Perricaudet M., Farrell P.J.;
RT "A spliced Epstein-Barr virus gene expressed in immortalized
RT lymphocytes is created by circularization of the linear viral
RT genome."
RL EMBO J. 7:769-774(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89095024; PubMed=2536113;
RA Sample J., Liebowitz D., Kieff E.D.;
RT "Two related Epstein-Barr virus membrane proteins are encoded by
RT separate genes."
RL J. Virol. 63:933-937(1989).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92219426; PubMed=1313931;
RA Busson P., McCoy R., Sadler R., Gilligan K., Tursz T., Raab-Traub N.;
RT "Consistent transcription of the Epstein-Barr virus LMP2 gene in
RT nasopharyngeal carcinoma."
RL J. Virol. 66:3257-3262(1992).
DR EMBL; M87778; AAA45885.1; -.
DR NON_TER 1
ET 1

```



[illegible]

DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE LATENT MEMBRANE PROTEIN 2A (FRAGMENT)  
GN LMP2.  
OS Human herpesvirus 4 (Epstein-Barr virus).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Lymphocryptovirus.  
OX NCBI\_TaxID=10376;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JIOYE AFRICAN BURKITT LYMPHOMA;  
RX MEDLINE=95146941; PubMed=7844523;  
RA Busson P., Hood Edwards R., Tursz T., Raab-Traub N.;  
RT "Sequence polymorphism in the Espdtein-Barr virus latent membrane  
protein (lmp)-2 gene."  
RL J. Gen. Virol. 76:139-145(1995).  
DR EMBL; X81771; CAA57375.1; -  
FT NON\_TER 118 118  
SQ SEQUENCE 118 AA; 12539 MW; 7B34420DE0B51373 CRC64;

Query Match 46.4%; Score 44.5; DB 12; Length 118;  
Best Local Similarity 50.0%; Pred. No. 5.9;  
Matches 9; Conservative 3; Mismatches 3; Indels 3; Gaps 1;

Qy 1 KDPX---GLEHDGNGXP 15  
:|||:||||:|  
Db 80 QDPSLYLGLQHDGNDGLP 97

## RESULT 14

O66556  
ID O66556 PRELIMINARY; PRT; 118 AA.  
AC O66556;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE LATENT MEMBRANE PROTEIN 2A (FRAGMENT)  
GN LMP2.  
OS Human herpesvirus 4 (Epstein-Barr virus).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Lymphocryptovirus.  
OX NCBI\_TaxID=10376;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=L2 NASOPHARYNGEAL CARCINOMA BIOPSY SAMPLE FROM CHINA;  
RX MEDLINE=95146941; PubMed=7844523;  
RA Busson P., Hood Edwards R., Tursz T., Raab-Traub N.;  
RT "Sequence polymorphism in the Espdtein-Barr virus latent membrane  
protein (lmp)-2 gene."  
RL J. Gen. Virol. 76:139-145(1995).  
DR EMBL; X81772; CAA57376.1; -  
FT NON\_TER 118 118  
SQ SEQUENCE 118 AA; 12461 MW; AB3A996BEA15ADFE CRC64;

Query Match 46.4%; Score 44.5; DB 12; Length 118;  
Best Local Similarity 50.0%; Pred. No. 5.9;  
Matches 9; Conservative 3; Mismatches 3; Indels 3; Gaps 1;

Qy 1 KDPX---GLEHDGNGXP 15  
:|||:||||:|  
Db 80 QDPSLYLGLQHDGNDGLP 97

## RESULT 15

O66557  
ID O66557 PRELIMINARY; PRT; 118 AA.  
AC O66557;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE LATENT MEMBRANE PROTEIN 2A (FRAGMENT)  
GN LMP2.  
OS Human herpesvirus 4 (Epstein-Barr virus).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Lymphocryptovirus.  
OX NCBI\_TaxID=10376;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JIOYE BURKITT LYMPHOMA FROM AFRICA;  
RX MEDLINE=95146941; PubMed=7844523;  
RA Busson P., Hood Edwards R., Tursz T., Raab-Traub N.;  
RT "Sequence polymorphism in the Espdtein-Barr virus latent membrane  
protein (lmp)-2 gene."  
RL J. Gen. Virol. 76:139-145(1995).  
DR EMBL; X81773; CAA57377.1; -  
FT NON\_TER 118 118  
SQ SEQUENCE 118 AA; 12480 MW; 9BD44A77EA15B979 CRC64;

Query Match 46.4%; Score 44.5; DB 12; Length 118;  
Best Local Similarity 50.0%; Pred. No. 5.9;  
Matches 9; Conservative 3; Mismatches 3; Indels 3; Gaps 1;

Qy 1 KDPX---GLEHDGNGXP 15  
:|||:||||:|  
Db 80 QDPSLYLGLQHDGNDGLP 97

Search completed: July 16, 2002, 11:21:59  
Job time: 302 sec

us-09-394-019a-248.rspt

Tue Jul 16 16:10:14 2002

**THIS PAGE BLANK (rspt)**